This is an Accepted Manuscript of an article published by Equinox in CALICO [The Computer Assisted Language Instruction Consortium] Journal in October 2020, available at https://journals.equinoxpub.com/CALICO/issue/view/3006 DOI: 10.1558/cj.40544.

Computer Assisted Language Instruction Consortium

calico journal (online) issn 2056–9017

Spotlight

Testing a Research-Based Digital Learning Tool: Chinese EFL Children's Linguistic Development

Hee Jin Bang[1], Kirsten Olander[2], and Erin Lenihan[3]

Abstract

This study examines the English language learning outcomes of young Chinese L1 children when using a digital self-guided learning tool, ABCmouse *English Language Learning Academy* (ABCmouse ELL), created by Age of Learning, Inc. and designed based on principles of second language research. In a pretest, post-test study design, experiment group participants used the digital learning tool to engage in English language learning activities for 15–20 minutes daily for six months while a control group used a comparable digital tool to do online math activities for an equivalent amount of time. Results showed that experiment group learners demonstrated significantly greater English language gains than control group peers. Data from parent surveys support these results, shedding light on children's overall engagement in the activities, and providing information about their motivation and confidence levels in English. Findings show that the research-based digital language learning activities in ABCmouse ELL were effective at promoting learning in this population, supporting the idea

Affiliations

[1]Age of Learning, Inc.
email: heejin.bang@aofl.com
101 N. Brand Blvd, 8th floor, Glendale CA 91203

[2]Age of Learning, Inc.
email: kirsten.olander@aofl.com

[3]Age of Learning, Inc.
email: erin.lenihan@aofl.com

©2020, equinox publishing

that teachers and parents can use tools like this to help younger learners develop English language skills in English as a Foreign Language contexts like China.

Keywords: foreign language learning; English language learning; digital language learning activities; young Chinese learners.

Introduction

Digital Game-Based Learning (DGBL)

Recently, there has been an increase in the number of digital activities and games designed to facilitate learning in various subject areas (e.g., Cobb & Horst, 2011; Kebritchi, Hirumi, & Bai, 2010; Kukulska-Hulme, Lee, & Norris, 2017; Reinhardt, 2017; Yang, Lin, & Chen, 2018). Research suggests that digital game-based learning (DGBL) is well suited for second language learning, particularly given its potential to lower anxiety and its ability to increase exposure to and use of the target language (Scholz, 2017; Sylvén & Sundqvist, 2017; Yang, Quadir, & Chen, 2015; Young et al., 2012). Yet, there are few pretest, posttest longitudinal studies examining the effectiveness of digital games in fostering language learning outcomes (Alyaz, Spaniel-Weiss, & Gursoy, 2017; Godwin-Jones, 2014). A review of the research (Hung, Yang, Hwang, Chu, & Wang, 2018) showed that studies focusing on very young children's use of digital games for foreign language learning are even more scarce. There is more research with older learners (e.g., De Wilde & Eyckmans, 2017; Scholz, 2017) which confirms that input through various media, including computer use and massively multiplayer online role-playing games, can be effective in language learning.

The existing research supports the idea that digital games can play an important role in young children's English language learning experiences (Jensen, 2017; Kukulska-Hulme et al., 2017), and when properly implemented in EFL contexts, such tools might motivate, engage, and spark the interest of young digital natives to learn English (Anyaegbu, Ting, & Li, 2012; Chiu, Kao, & Reynolds, 2012). Recent studies suggest that the game design strongly affects the effectiveness of DGBL (Benton, Vasalou, Barendregt, Bunting, & Révész, 2019; Berkling, & Gilabert Guerrero, 2019), and that learner engagement is critical to producing the desired learning outcomes (Chen, 2018). The digital tool tested in the current study was designed in accordance with such research. We turn to these principles next.

Developing a Digital Tool ELL App Based on Principles from SLA Research

A team comprising curriculum specialists, English teachers, second language acquisition (SLA) researchers, applied linguists, and game developers collaborated to create ABCmouse ELL, a digital L2 teaching and learning (L2TL) program that is developmentally appropriate for young children learning English in a foreign language context. Since SLA research shows that learners learn best through activities or tasks that use language as a vehicle to accomplish a goal, rather than those designed purely for the purposes of learning the grammatical aspects of the language (e.g., Heift, Mackey, & Smith, 2019; Long, 1985, 2016), the team created the program around English language activities and experiences that would engage young learners in authentic contexts (i.e., contexts relevant and/or familiar to children based on their lives outside the language learning tool) (Long, 2015, 2016). ABCmouse ELL was designed to provide immersive, engaging learning experiences that involve familiar topics, like food, pets, games, and toys, and involved learners interacting with the activities, constructing knowledge through trial and error, and producing and using the target language (DeKeyser, 2007), with opportunities to receive the right amount of developmentally appropriate implicit and explicit corrective feedback at the right time (Carroll & Swain, 1993; Mackey, 2012). It was designed to provide rich, meaningful input, incorporating the target language through pedagogical tasks and content (e.g., videos of conversations, songs), in ways that would be comprehensible at the learners' proficiency levels, including activities that build on the learners' developing linguistic knowledge (Gass, 2017). Both the Common European Framework of Reference (CEFR) standards and the American Council on the Teaching of Foreign Languages (ACTFL) guidelines were used, in conjunction with second language research findings, to underpin the developmental sequences in which language is presented and produced, meaning that language is provided to learners at points where research has shown to be developmentally appropriate and learnable (Pienemann, 1998).

ABCmouse ELL also incorporates principles from spaced repetition theory, which has recently been incorporated into cutting-edge SLA research (Serrano & Huang, 2018). This idea posits that the timing of review and practice affects learning and that practice can be more effective when spaced out over time rather than being grouped together (Kang, 2016; Nakata & Suzuki, 2019; Rogers, 2017). Based on research into techniques for efficient encoding into memory, the activities in ABCmouse ELL repeat and review content in an algorithm-driven pattern to improve long-term retention (Tabibian et al., 2019). ABCmouse ELL also incorporates empirically supported effective practices from second language instruction, such as scaffolded target language input (Moeller & Roberts, 2013). The emphasis on communication in the activities

combines presentation, practice, and production of skills, with interactive activities emphasizing the development of fluency (Shrum & Glisan, 2009).

Making Digital Activities Effective: Linguistic and Language Considerations

The linguistic input in ABCmouse ELL, in terms of its comprehensibility and opportunities for production, was also driven by research into language, education, and learning. ABCmouse ELL is designed so that young children learn to understand and use vocabulary and simple grammar. Games are a useful part of the vocabulary exposure and learning process because they are motivating and challenging, providing learners with opportunities to hear and practice using language in non-stressful contexts (Uberman, 1998). The utility of games can be seen, for example, in a study conducted with 11–12 year-old ESL learners in Montreal by Cobb and Horst (2011), who found that using a suite of vocabulary training games for two months was associated with the kinds of gains in vocabulary recognition that they claim are normally achieved in one to two years. Another study conducted with English language learners (ELLs) who were 8 and 10 years old in Denmark found that gaming with oral and written English input was significantly related to children's vocabulary knowledge (Jensen, 2017). A substantial body of literature indicates that games built on tasks that are fun and meaningful for the users are more effective language learning tools than games designed to practice discrete grammar or decontextualized vocabulary (Chen, Tseng, & Hsiao, 2016). For these reasons, ABCmouse ELL includes engaging activities that introduce young learners to language in both receptive and productive contexts that are meaningful and enjoyable for learners.

For grammar to be effectively internalized and automatized, studies have suggested that learners should have opportunities to attend to linguistic form as part of their efforts to engage with meaning (in contrast with attending to linguistic forms, which implies focusing attention on forms of the target language detached from any meaningful communicative context) as summarized in Doughty and Williams (1998). ABCmouse ELL was designed to introduce grammar ("form") embedded in task-based game contexts that typically involve communication amongst characters in the game. Finally, opportunities for developmentally appropriate corrective feedback that helps learners understand their errors is also embedded in ABCmouse ELL so that learners receive timely correction throughout their learning process (Dean, Hubbell, Pitler, & Stone, 2012; Mackey, 2012).

Keeping Learners Motivated and Engaged

Research into second language motivation also underpins the design of the program. Motivation is widely accepted as a contributing factor in second language learning progress, and it is now understood as being dynamic and malleable (for a review, see Csizér, 2017). Motivation can also fluctuate depending on several variables, including the language activity or task the learner is carrying out, their level of interest or engagement, and their interlocutor in an instructional setting. Motivation researchers such as Dörnyei (2014), Mercer (2020), and Lasagabaster, Doiz, and Sierra (2014) have recently synthesized many strands of motivational research in describing a comprehensive construct with explanatory power—direct motivational currents (DMCs). These are enjoyable periods of high involvement and engagement in a task or activity that propel learners towards a highly-valued goal or end point. A great deal of empirical research has linked motivation and L2 learning outcomes (Henry, Davydenko, & Dörnyei, 2015; Masgoret & Gardner, 2003). Given such evidence, ABCmouse ELL was designed with the principle of promoting learner engagement and motivation at the forefront.

Additionally, ABCmouse ELL aims to foster independent, active learning. A learning path in the game guides children along a sequence of lessons with gradually more challenging communicative objectives, and the program is self-guided in the sense that learners can choose content to explore based on their interests from a rich resource of library options that include music, paintings, games, and books on a variety of topics. The program also includes areas (i.e., personal rooms, shops) where children can use tickets (earned by completing activities in the game) to purchase pets, clothing, furniture, and so on to personalize their avatars and rooms. In sum, ABCmouse ELL aims to help young learners develop autonomy as they make choices about what to explore and how to shape their own learning environment in the app.

In summary, core tenets of SLA research and applied linguistics informed the construction of the ABCmouse ELL's narrative-rich input, customizable features, and in-app interactive activities. We now turn to the context for which ABCmouse ELL was designed: young children learning English in China.

English as a Foreign Language Education in China

In 1978, the Chinese Ministry of Education (MOE) made English a compulsory subject starting in the third year of primary school, but Chinese children are now starting to learn English at younger ages (Wenting, 2019). Chinese policymakers recently transformed the English language curriculum to focus on the development of communicative competence (Wang, 2009) to better prepare children for the 21st-century global economy in which English proficiency is

essential (Hu, 2005). While these reforms produced some positive outcomes (e.g., innovative approaches to teacher development, greater teacher autonomy in terms of curriculum), there is still a documented list of problems (Feng, 2006), including limited availability of English print materials necessary to develop literacy skills in English and a scarcity of qualified English language teachers in China (Zhang, 2012).

Updated English curriculum standards in China issued in 2011 emphasized task-based learning and provided more specific guidelines for English literacy education for grades 3–6 (ages 8–12) (Chinese Ministry of Education, 2011). However, these standards do not require introducing English language instruction in grades 1 and 2 (ages 6–7), meaning there is a lack of guidance and standardiza- tion for English education for younger children. Media reports suggest parents and EFL teachers in China are very interested in finding effective educational materials and resources designed to help their children develop communica- tive skills in English from early on.

ABCmouse ELL was developed to meet these needs in the form of a digital tool based on principles from SLA research, designed for the Chinese context, to be used by very young children. The question explored in this study is: to what extent can a research-driven app such as ABCmouse ELL help young learners in an EFL context develop English language skills?

Methods

The study was approved by a US-based Institutional Review board, and the researchers obtained informed consent from the parents of all participants in the study.[1] A paid team of design and research consultants in China helped with recruitment, communication with participants, and logistics associated with data collection. Two native Mandarin speakers (with experience of teaching English to young learners) were trained to assess the children's language skills. They were blind to the research question and to experiment or control conditions of the children, and the order in which children were assessed was randomized.

Participants

Children were recruited from Hangzhou, the fourth largest metropolitan area in China, in fall, 2018. The children were recruited from 88 different kindergartens in the area. To be eligible for the study, the children had to be between 5 and 6 years old at the start of the study, with little to no prior English knowledge (operationalized as knowing fewer than 20 English words). They needed to have access to a smartphone or tablet for up to 20 minutes a day, and parents

needed to commit to a six-month study period. Children whose parents had strong English skills, or who were professionals in industries with a substantial amount of English usage (e.g., English teaching, marketing) were ineligible for participation. There were no significant differences between groups in terms of parents' educational levels (62.1% of control and 60.7% of experiment group parents reported having completed junior college) or income (51.7% of control versus 50.8% of experiment group parents reported monthly income of 1,423–2,846 USD).

Experiment and Control Conditions

The experiment group children were asked to use ABCmouse ELL, the digital English Language program, for a minimum of 15–20 minutes daily for six days a week between December 2018 and May 2019. Researchers monitored the usage data on a weekly basis via surveys completed by parents and by reviewing the game usage data, and efforts were made to ensure all participants met the minimum threshold for usage each week (e.g., communications with parents, small gifts such as school supplies for those who made the greatest improvement in terms of making up for missed time). The control group children used a digital math learning app for the same amount of time over the six months. Researchers decided to have the control group use a math app rather than a comparable language learning app because research has shown that children can often gain some basic level of English knowledge through environmental exposure (e.g., games, television, Internet; see De Wilde & Eyckmans, 2017). Requesting that the control group children spend the same amount of time engaged in technology-based activities that involve no English was an effort to provide similar learning conditions for both groups and to facilitate the examination of the children's engagement with ABCmouse ELL, while controlling for any extra-experimental effects of informal English learning that can occur through everyday activities. None of the participants in the analytic sample enrolled in any other English classes or studied English outside of the app for the duration of the study.

Experimental Materials

The child participants used the first two levels of the program, which targeted a total of 165 words and 15 simple sentence structures. The themes covered in these levels are developmentally appropriate and familiar to learners, and include animals, common objects, food, places, colors, numbers 1–15, shapes, people, body parts, clothing, and classrooms. Table 1 displays the communicative objectives by language skill across these levels, and Figures 1–2 show screenshots of sample games that students played in the program.

Table 1
Linguistic Objectives of the First Two Levels of the Digital Learning Activities

Listening Comprehension	Speaking	Vocabulary and Sentence Structure
Recognize familiar, everyday words delivered in clearly defined, everyday contexts	Produce short phrases about themselves, giving basic personal information	Have a basic vocabulary repertoire of words and phrases related to concrete situations
Understand short, simple questions and statements delivered clearly and accompanied by visuals and repeated if necessary	Describe themselves (e.g., name, age, family) using simple words and expressions	Use very simple principles of word order in short statements
Answer yes/no and wh-questions and respond using gestures, words, and simple phrases	Say how they are feeling using simple words (e.g., happy) accompanied by body language	Use frequently used verbs (e.g., walk, eat) in guided activities
Understand short, simple instructions for actions (e.g., stop) delivered face-to-face, accompanied by pictures and repeated if necessary	Point to something and ask what it is	Expand noun phrases in simple ways to enrich sentence meaning and add details in guided activities
	Use isolated words and basic expressions	
	Manage very short, isolated utterances using gestures	Expand sentences with frequently used prepositional phrases (e.g., in the house) in guided activities

Figure 1. Bubble Popper game: player hears the word "fish." The task is to tap on all of the bubbles containing fish and ignore the bubbles containing distractors.

Control Materials

To control for (a) time spent online by young children playing an educational game, and (b) any extra-experimental environmental input, a comparable math app was selected. The control children were asked to spend 15–20 minutes a day, six days a week, playing games in this math app designed to improve their math skills.

Research Design

After balancing pretest scores, age, and gender, a total of 66 children were randomly assigned to the experiment group and 56 children to the control group. A total of nine children (three experiment, six control) were subsequently excluded from the analyses because six (one experiment, five control) enrolled in an English class; two experiment group participants were unable to attend an assessment; and one control participant did not spend enough time using the math app. The final analytic sample was 113 (63 experiment, 50 control).

Pre- and Posttests

Two measures of English language proficiency, one external and one internal, were used to assess children's language skills before and after the intervention. The external measure was the IDEA Language Proficiency Test (IPT) (Ballard & Tighe, Publishers, 2019), a widely-used assessment of overall English language skills found to be accurate and reliable in studies involving speakers of Spanish, Chinese, Japanese, and Korean (Cook, 1995; Stansfield, 1991). The

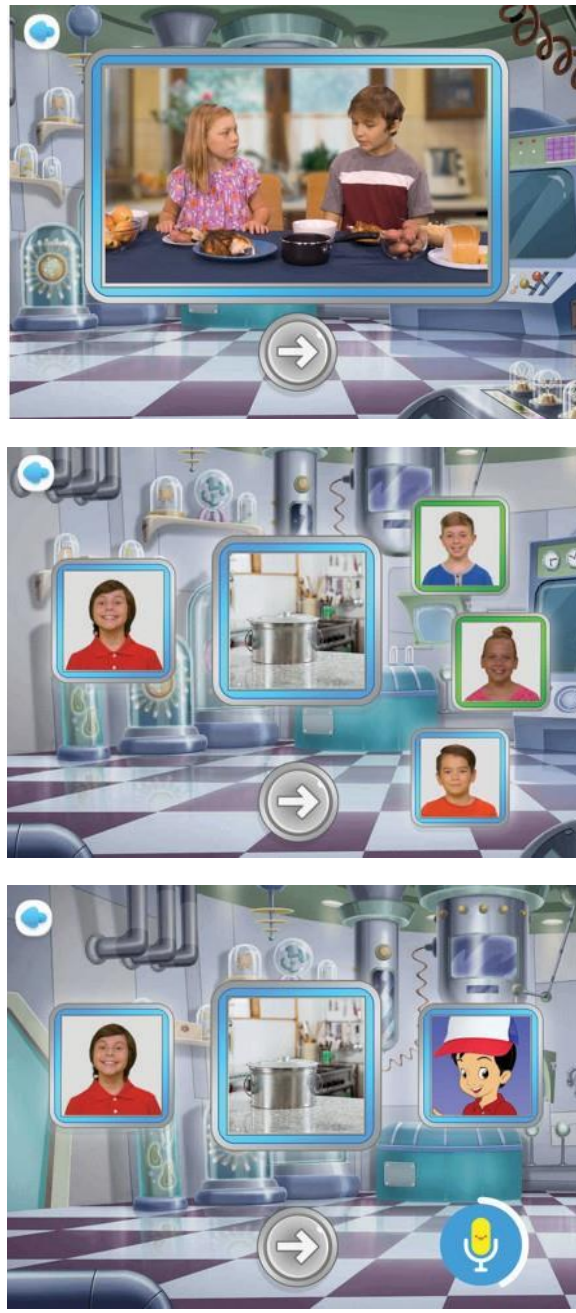

Figure 2. Talk Time game. a) Player hears a conversation between two children. b) Player selects a correct answer from three sentences produced by three children. c) Player is asked to produce an answer to the question (microphone).

Pre-IPT Oral Proficiency Test was selected because it was designed for very young learners. The examiner asked the questions in English using a story-board and cardboard pieces, and a total of 10 questions were administered, targeting vocabulary, grammar, comprehension, and language functions. For each correctly answered question, one point was awarded. A sample item is in Appendix A.

An internal test was also used. This was created by the researchers, using linguistic forms that children did not know at the beginning of the study but were exposed to through the game during the study, in terms of both comprehension and production. This Internal English Proficiency Test included a total of 25 questions divided into five subsections (five questions per section, see Table 2). For each question in the assessment, the test administrator gave the child up to 20 seconds to provide the correct answer and did not provide feedback to indicate whether the child produced the correct answer. Partial credit was given in some cases (e.g., producing a single word answer rather than a complete sentence or selecting a picture of three pencils instead of three ducks). Sample items from each of these subsections are in Appendix A.

Table 2
Description of Internal Assessment Subsections

Section	Task	Total Points
Vocabulary Identification	Listen to the audio of a target word and select an image that matches the word from a collection of 3 images and a question mark (indicating "I don't know").	10
Listening for Meaning	Listen to 1-2 sentences and identify a picture corresponding to the meaning of a sentence(s) from a collection of 3 images and a question mark.	30
Speech Production	Look at a picture, listen to a question about the picture ("What do you see?"), and respond with an appropriate answer.	20
Conversation	Listen to a personal question (e.g., How old are you?), and respond with an appropriate answer.	30
Pronunciation	Listen to individual words and repeat each word.	10

Parent Surveys

A brief written survey (in Mandarin) was administered to the parents at the end of each week during the study to collect information on program usage. The survey questions included parental ratings of the extent to which children appeared engaged while using the program and any changes parents perceived in their children's English abilities. An end-of-study survey was also administered to parents to gather more in-depth information about their children's experiences with the digital program. The questions asked about the degree to which children made improvements in various English language skills and the impact that the program had on their children as learners. Focus group data were also collected. Figure 3 provides an overview of the research design and data collection activities.

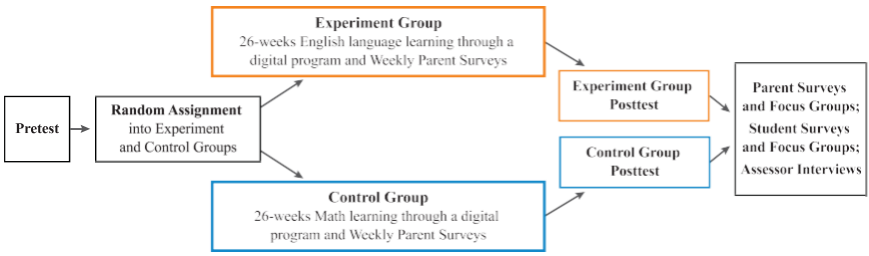
Experiment Group
26-weeks English language learning through a digital program and Weekly Parent Surveys
Pretest
Random Assignment into Experiment and Control Groups
Control Group
26-weeks Math learning through a digital program and Weekly Parent Surveys
Experiment Group Posttest
Control Group Posttest
Parent Surveys and Focus Groups; Student Surveys and Focus Groups; Assessor Interviews

Figure 3. Overview of study design and procedures (orange = experiment group; blue = control group; black = both groups).

Interrater Reliability

The Pronunciation scores were the only data that involved judgements. Two trained bilingual raters who had not administered the tests and were blind to experiment versus control conditions scored all the audio files. Their scores were compared using simple percentage agreements. The initial interrater agreement rate was 78%. Discussion and coding socialization was practiced whereby both raters listened to the disagreed-upon data together. Through this process, 100% agreement rate was achieved.

Analysis and Results

As noted above, the research question addressed was: to what extent can a research-driven app such as ABCmouse ELL help young learners in an EFL context develop English language skills?

Overall English Language Proficiency

To create an initial overview of the children's English language skills before and after using the programs, the two tests (IPT and Internal English Proficiency Test) were combined to create an overall English Language Proficiency (ELP) score from a weighted average of the total number of items in each test (10 items in the IPT and 25 items in the Internal English Proficiency Test).

Table 3 shows the average raw scores and standard deviations on the overall ELP, along with the results of an independent-samples t-test comparing group means. A visual representation of these scores is shown in Figure 4. At pretest, the control and experiment group students performed similarly, with no statistically significant differences between the two groups' scores. However, by posttest, the experiment group children significantly outperformed their control group peers. The effect size was 2.17 (Cohen's *d*), indicating that the language learning app was effective in improving young children's overall ELP.

Table 3
Control and Experiment Group Scores on Overall English Language Proficiency Pre- and Posttest

	Pre Mean (*SD*)	**Post Mean (*SD*)**
Control	15.74 (8.67)	17.54 (10.57)
Experiment	16.21 (8.45)	47.65 (15.73)
t-test	t (111) = .77	t (111) = 11.61***

*** $p < .001$
Note: A repeated measures ANOVA with time and group indicated a similar pattern of results.

An examination of the relationship between the amount of app usage and children's performance on the overall posttest showed a strong positive correlation (Plonsky & Oswald, 2014) ($r = .65$, $p < .001$). An examination of the total amount of time children used their assigned programs throughout the experiment showed that the experiment group on average used ABCmouse ELL for 29.29 hours (SD = 8.21), more than double the amount of time that control group children used their math app (11.99 hours, SD = 2.32). Given the disparity between control and experiment groups in the usage time of their respective apps, we examined the relationship between usage and performance on the posttest by group, dividing each group into three usage groups (low-, medium-, and high-usage, Table 4). A graph of the posttest scores for each subgroup (not displayed here) suggested that experiment group children's

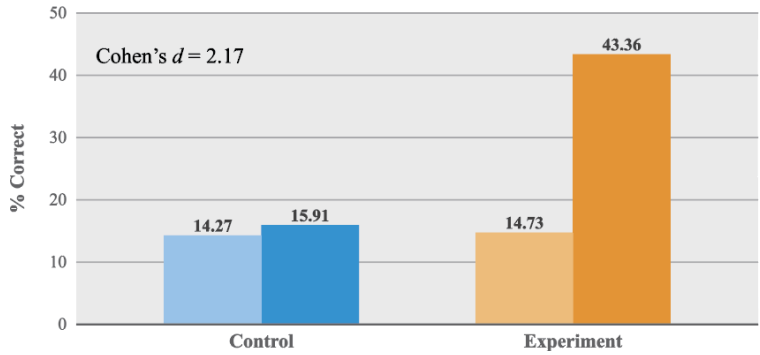

Figure 4. Overall pre- and post-assessment English Language Proficiency scores for control (blue) and experiment (orange) groups.
Note: Lighter shades of blue and orange represent pretest scores; darker shades represent posttest scores.

performance on the posttest is largely driven by their degree of engagement with ABCmouse ELL.

Table 4
Control and Experiment Group's Low-, Med-, and High-Usage hours

	Control	Experiment
Low	6.50-10.91 hours (n=16)	9.85-24.78 hours (n=21)
Med	10.92-13.02 hours (n=17)	24.79-32.29 hours (n=21)
High	13.03-16.40 hours (n=17)	32.29-50.27 hours (n=21)

Idea Proficiency Test and Internal English Proficiency Test

When the results of the IPT Assessment and the Internal English Proficiency Test were examined separately, no significant differences were observed at pretest between control and experiment groups, but the differences between the two groups were significant at posttest. As shown in Table 5 and Figure 5, the experiment group children scored about three times as high as their control group peers on each of the two measures.

Table 5
Control and Experiment Group Scores on IPT and Internal English Proficiency Test at Pretest and Posttest

	IPT Mean (*SD*)		Internal English Proficiency Test Mean (*SD*)	
	Pre	**Post**	**Pre**	**Post**
Control	.86 (.90)	.84 (.74)	14.88 (8.57)	16.70 (10.40)
Experiment	.73 (.68)	2.37 (1.13)	15.48 (8.17)	45.29 (15.02)
t-test	*t* (111) = .87	*t* (111) = 8.27***	*t* (111) =.38	*t* (111) = 11.45***

*** $p < .001$

While the effect size on the IPT assessment of 1.56 is lower than that of the Internal English Proficiency Test (2.15), it is noteworthy that the external validated assessment (i.e., not aligned with the syllabus of linguistic forms underlying the app) showed the same significant patterns.

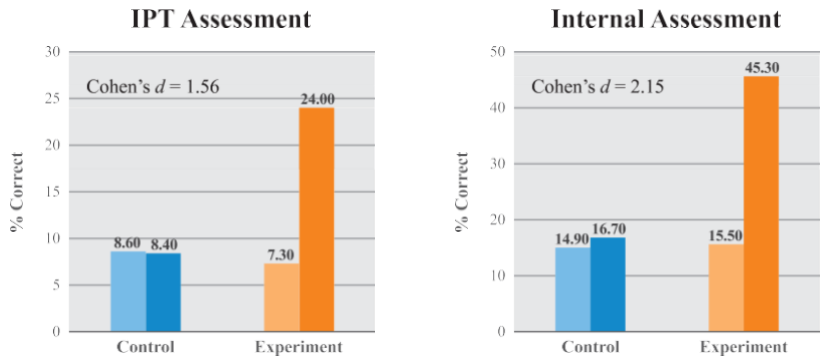

Figure 5. Pre- and posttest performance on the IPT and Internal English Proficiency Test.

Internal English Proficiency Test: Linguistic Analysis

The pre- and posttest comparisons of individual subsections showed that experiment group children made statistically significant improvements on each of the language skills assessed. As shown in Figure 6, the overview of the learners' performance on the five internal assessment subsections shows

experiment group children scoring substantially higher than their control group peers on each of the subsections.

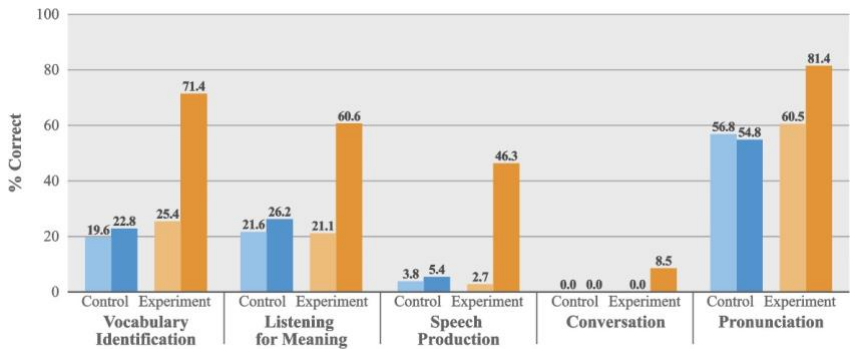

Figure 6. Pre- and posttest performances across the internal assessment sections.

Vocabulary Skills

Table 6 displays the raw mean scores and standard deviations, along with the results of independent samples t-tests for the Vocabulary Identification and Listening for Meaning sections of the internal assessment. Each child listened to the audio of a target word or sentences and selected an image matching the word or sentences from a collection of three images and a question mark (i.e., "I don't know"). Since four answer choices were available for each question, children could guess and manage to answer correctly one out of four times (i.e., 25% of the time). Figure 6 shows the scores in terms of percentage correct, with the pretest scores for both groups being close to the result that would be

Table 6
Control and Experiment Group Scores on *Vocabulary Identification* and *Listening for Meaning*

	Vocabulary Identification		**Listening for Meaning**	
	Pre Mean (*SD*)	**Post Mean (*SD*)**	**Pre Mean (*SD*)**	**Post Mean (*SD*)**
Control	1.96 (2.27)	2.28 (2.25)	6.48 (6.13)	7.86 (7.32)
Experiment	2.54 (2.16)	7.14 (2.18)	6.33 (6.12)	18.19 (7.02)
t-test	t (111) = 1.39	t (111) = 11.62***	t (111) = .13	t (111) = 7.62***

*** $p < .001$

expected from participants' guessing at the answers. By posttest, the control group children's performance remained the same, at chance level, while the experiment group children made significant gains. They correctly identified nearly 4 out of the 5 words in the Vocabulary Identification section and 3 out of the 5 sentences in the Listening for Meaning section. The effect size (Cohen's *d*) of 2.18 and 1.43 in the two sections, respectively, provided clear evidence that the activities were effective in terms of children's learning of vocabulary words.

Speaking Skills

The raw mean scores, standard deviations, and t-test results for the Speech Production and Conversation sections of the Internal English Proficiency Test are shown in Table 7. In Speech Production, which required the child to look at a picture, listen to a question, such as "What do you see?", and respond with a word representing the picture, children did not have the option of guessing, as evidenced by the scores close to zero for both groups at pretest and for the control group children at posttest. In the Conversation section, each child listened to personal questions such as "What color do you like?" and responded with an appropriate answer. At pretest, none of the children were able to engage in any conversation in English, and at posttest, the control group children continued to have no English conversation skills. By posttest, children who used ABCmouse ELL answered on average about 3 out of the 5 Speech Production questions correctly (effect size = 1.8) and 18 children (29% of the experiment group) received at least partial credit for one of the five questions in the Conversation section (effect size = 0.68).

Table 7
Control and Experiment Group Scores on *Speech Production* and *Conversation*

	Speech Production		**Conversation**	
	Pre Mean (*SD*)	**Post Mean (*SD*)**	**Pre Mean (*SD*)**	**Post Mean (*SD*)**
Control	.76 (1.89)	1.08 (2.49)	.00 (0)	0 (0)
Experiment	.54 (1.84)	9.25 (5.60)	.00 (0)	2.56 (4.99)
t-test	*t* (111) = .63	*t* (111) = 9.59***	n/a	t (111) = 3.62***

*** $p < .001$

Pronunciation

The Pronunciation section required each child to listen to audio recordings of each target word and repeat them. Five target words ("pig," "book," "man,"

"bed," and "cookie") were selected taking into account the phonemes that students had exposure to in the program as well as the phonemes that were developmentally appropriate for kindergarten students (Gillon, 2004; Paulson, 2004). As shown in Table 8, children performed best on this section of the pretest, possibly a reflection of the relatively easy task of repeating what is played on an audio file, without the need to understand the meaning of the target words. The posttest results show that while control group children made no gains, those who used ABCmouse ELL showed substantial improvements in pronunciation, producing "good" pronunciations of four out of the five words in this section (effect size = 1.2). For example, for the word "pig," a "good" pronunciation (worth two points) was one in which all phonemes were produced correctly; a "fair" pronunciation (worth one point) had one mispronounced phoneme (e.g., /pɪt/, /pig/, /pɪk/, /peɪg/, /bɪg/); and a "poor" pronunciation (worth zero points) had two or more mispronounced phonemes and/or was incomprehensible (e.g., /dɛk/).

Table 8
Control and Experiment Group Scores on ~~Pre- and Posttest:~~ *Pronunciation*

	Pre Mean (*SD*)	**Post Mean (*SD*)**
Control	5.68 (2.59)	5.48 (2.35)
Experiment	6.05 (2.59)	8.14 (2.06)
t-test	t (111) = .46	t (111) = 6.41***

*** $p < .001$

To examine the extent to which ABCmouse ELL had an impact on the pronunciation skills of children who did not obtain "good" scores on their pronunciations of the target words, we excluded students whose production was rated "good" at pretest on the basis of a ceiling effect. Across the five test words, between 71.4% to 82.1% of experiment students who produced "fair" or "poor" pronunciations at pretest made improvements by posttest. The percentage of students who moved from "poor" to "good" ranged from 24.2 to 43.6, while the percentage of students who moved from "fair" to "good" ranged from 26.9 to 48.5. In other words, children who used the language learning program demonstrated notable improvements in their pronunciations of all five words assessed.

Survey Results

The assessment results were triangulated with feedback collected through parent surveys, which enabled us to assess parents' subjective impressions and the aspects of ABCmouse ELL that were designed to promote engagement and motivation. The majority of experiment group parents confirmed the quantitative findings, reporting that their children made improvements in speaking (87%), understanding (75%), and reading (81%) English words or sentences. They also reported that they believed ABCmouse ELL helped their children become more interested (75%), motivated (71%), and confident (79%) in English. All of these factors have been implicated in developing fluency in second language learning (Zheng, Young, Brewer, & Wagner, 2009).

Additionally, over the course of the study, 74% of the experiment group parents reported on the weekly surveys that their children were "always" or "very frequently" engaged while using ABCmouse ELL. Nearly 60% of the experiment group parents indicated that their children voluntarily produced English words and phrases outside of the assigned program usage time at least several times a week. More information on the qualitative data that provided helpful in-depth context for these survey responses, including focus groups, are the topic of a future publication on the efficacy of ABCmouse ELL at promoting engagement.

Discussion

The results of this study show that after using a digital language learning program for 15–20 minutes a day over a 6-month time period, five- and six-year-old Chinese children demonstrated substantial improvements in both comprehension and production skills in English. These findings indicate that consistent usage of a digital English learning tool designed based on principles of research on SLA can be highly effective in helping young learners in an EFL context build their English language skills. Given that these children came from 88 different kindergartens and care was taken to ensure that they were representative of average five- to six-year-old children learning English in China in that region, these results may be generalizable to other Chinese children with similar demographic features in similar cities.

The study results also add to the existing body of research on DGBL which illustrates that digital activities are helpful in promoting language learning (Yang et al., 2015; Young et al., 2012) while also filling a gap in the literature by focusing on digital game use by very young children learning English in a foreign language context. The learning outcomes, backed up by the parental surveys, indicate that these children made substantial linguistic gains, and also enjoyed their time with the program. The fact that the majority of experiment

children were engaged while using ABCmouse ELL further corroborates the existing research on the importance of motivation and engagement in L2 learning.

Additionally, the results are consistent with the general understanding in SLA that learners generally start by first developing receptive skills (i.e., listening, reading), the knowledge of which slowly transfers to productive skills (i.e., speaking, writing). We saw that the results of the Vocabulary Identification and Listening for Meaning sections on the internal assessment showed greater improvements than the results of the Speech Production and Conversation sections. The fact that the program appeared to have the strongest impact on children's vocabulary acquisition is both unsurprising and promising given that building learners' vocabulary knowledge is an important phase in language acquisition, the foundation upon which both receptive and productive skills are built (Golkova & Hubackova, 2014; Zhou, 2010).

Based on these empirical findings, and in the absence of guidance or standardization for English education from the Chinese Ministry of Education in grades 1 and 2, we believe the particular game-based approach tested here offers an effective way for EFL teachers and parents to help young Chinese children develop communicative skills in English, and/or to supplement English instruction for students at schools where there may be a shortage of qualified English language teachers with experience implementing communicative and task-based approaches to language instruction. Our findings might not apply to all DGBL programs, however, particularly those not designed based on SLA principles.The parental survey results suggest that ABCmouse ELL was an effective resource to help children get an early start in English. The native English speakers featured in the activities provided the Chinese children with opportunities to hear and emulate accurate pronunciations of English words. The design ofthe game, emphasizing authentic communication, was particularly valuable inEFL contexts where limited input in the target language is available in learners' everyday lives. Such programs may be a cost-effective learning resourcefor parents who may not have the resources or the time to take their childrento English language classes, but who, nevertheless want their children to have an early start in English (Chen, 2018).

While the current study is small in scope, work on distance, self-paced, and digital game-based learning for children is increasingly relevant, as educators evaluate the utility of diverse learning settings, particularly for very young children still developing L1 proficiency in a context such as China where there is relatively little exposure to English. Furthermore, interest in technology-enhanced language learning seems likely to increase in the future, given general advances in technology and potential changes to education like the ones

occurring at the time of writing in 2020—a time of pandemic and physical distancing.

Limitations and Directions for Future Research

Overall, this study indicated that ABCmouse ELL was effective. As briefly discussed above, it would be advisable and interesting for future studies to compare different kinds of programs designed to teach language (e.g., comparisons of ABCmouse ELL with Duolingo Kids, Lingo Kids, Monkey Junior, etc.), as opposed to comparing a language program with a non-English language learning program designed to teach children math skills. While the control group's app was not quite comparable to ABCmouse ELL, we made this comparison to rule out environmental or extra-experimental English input. Our findings, therefore, shed light only on ABCmouse ELL, not on other programs designed to teach children English language skills.

Future studies including comparison groups using other English or language learning programs will enhance our understanding of the extent to which such programs are effective in helping young learners develop language skills. Our research, however, has clearly shown that the research-based ABCmouse ELL was effective at doing what it set out to do, which is developing English skills. Notably, we tested it with younger learners across a longer period of time than is typically the case in applied linguistics and CALL research.

Future research could also include more detailed analysis, comparing this program with different types of language learning activities and games to examine how DGBL fosters young learners' language vocabulary, listening comprehension, speech production, and pronunciation skills, as well as the extent to which they produce learning gains on various skills. It would also be interesting to compare situations and contexts where the program is used in isolation and in combination with instruction, or other synchronous or asynchronous technological learning methods. As many learners engage in combinations of in-person and online instruction, such research may become increasingly important.

Game design for DGBL programs for children is an area ripe for more investigation. For example, further research could explore how children's motivation for learning is affected by game incentives and focus on learning outcomes like preparation for future learning, introduction of new knowledge, and practicing of already introduced knowledge as a means of evaluating the success of a learning app, rather than focusing on comparisons of control and experimental group gains (Berkling & Gilabert Guerrero, 2019). Further attention should also be paid to how breakdowns in game design can impede learning and how

these breakdowns can be turned into breakthroughs through instructional or game components (Benton et al., 2019).

The introduction of programs like this one into schools in China could lead to future research examining how teachers integrate the program into their instruction and the efficacy of the program in producing or enhancing gains in students' English language skills. An investigation of how teachers use the program to supplement their instruction or in ways which have come to be known as "flipped classrooms" would provide valuable insights into the features that optimize instruction time. Such studies may also generate recommendations for content modification to make them more culturally appropriate and relevant.

Conclusion

The promising results from this study indicate that young children can learn important skills in English by using ABCmouse ELL. The program's focus on listening and speaking skills through themes and topics that are common in everyday life is designed to help children acquire English as a tool for communication. Moreover, it is noteworthy that the program fostered children's interest, motivation, and confidence in learning English, all of which are precursors to achievement in English skills, including good pronunciation. Future studies will be essential in not only guiding future development and refinement of this program, but also in deepening our understanding of how digital activities can effectively teach languages and foster engagement simultaneously.

Notes

1. This study was approved by Solutions IRB (https://www.solutionsirb.com), an independent fee-paying institutional review board in the United States, which is fully accredited by the non-profit Consortium of Independent Institutional Review Boards. The Office for Human Research Protections (OHR)-assigned registration number for Solutions IRB is 00008523, and their IORG # is IORG0007116. The IRB approval number for this study is: #2018/11/11.

Acknowledgments

The authors would like to thank Jacey Chen, Phoebe Jiang, Paula Hidalgo, Jerry Chen, Yuying Mesaros, Zac Katz, employees of Age of Learning, who playedsignificant roles in developing the curriculum and instruments for the study and providing support for various aspects of the study. Thanks are due to a

number of paid consultants to the project, especially Alison Mackey (Georgetown University), a member of the Age of Learning Curriculum Advisory Board who contributed to the literature review, background, discussion, and write-up, along with Nicole Ziegler (University of Hawai'i) and Kendall King (University of Minnesota), who read and provided helpful comments on an earlier draft of the paper. The authors would also like to thank Chao Su and his colleagues at One Leap, the paid team of research and design consultants in China who helped with recruitment and data collection. Final thanks go to three anonymous CALICO reviewers and of course, to the editors of the CALICO Journal for their invaluable feedback throughout.

About the Authors

Hee Jin Bang, Kirsten Olander, and Erin Lenihan are salaried employees of Age of Learning, Inc. who contributed to various aspects of the research, including design, planning, management of the data collection process, coding, analysis, as well as the write-up.

Hee Jin Bang, PhD, serves as the Director of Efficacy Research and Evaluation and acted as lead author throughout, coordinating all aspects of the paper, including the review and revision process. Dr. Bang's areas of expertise include evaluation of educational programs, second language acquisition, and education of English Language Learners. She holds a PhD in Education (New York University), MEd in Human Development and Psychology (Harvard University), and BA Honors in Linguistics and French (University of Oxford).

Kirsten Olander, Research Assistant for Efficacy Research and Evaluation, coordinated all communication with partners in China and provided invaluable support in data management and literature review. Kirsten received her BA in Psychology (UCLA) and is planning to pursue her doctorate in Psychology focusing on environmental factors that shape resilience in children.

Erin Lenihan, Senior Director of ELL Curriculum and Design led the development of the ABCmouse ELL curriculum, bringing over a decade of experience in international teaching, curriculum development, and teacher training. Erin holds an MSE in Curriculum and Instruction: Teaching English to Speakers of Other Languages (TESOL) (University of Kansas) and Teaching English as a Foreign Language (TEFL) certification (UC Irvine).

References

Alyaz, Y., Spaniel-Weiss, D., & Gursoy, E. (2017). A study on using serious games in teaching German as a foreign language. *Journal of Education and Learning, 6*(3), 250–264.

Anyaegbu, R., Ting, W., & Li, Y. I. (2012). Serious game motivation in an EFL classroom in Chinese primary school. *The Turkish Online Journal of Educational Technology*, *11*(1), 154–164.

Ballard & Tighe, Publishers (2019). Idea Proficiency Test. http://www.ballard-tighe.com/home/

Benton, L., Vasalou, A., Barendregt, W., Bunting, L., & Révész, A. (2019). What's missing: The role of instructional design in children's games-based learning. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. New York, NY: Association for Computing Machinery.

Berkling, K., & Gilabert Guerrero, R. (2019). Designing a comprehensive evaluation method for learning games: A general approach with specific application to iRead. In J. Koivisto & J. Hamari (Eds.), *Proceedings of the 3rd international GamiFIN conference*.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, *15*(3), 357–386.

Chen, L. (2018). Chinese parents spend up to US$43,500 a year on after-school classes for their children. *South China Morning Post*. Retrieved from https://www.scmp.com/news/china/society/article/2176377/chinese-parents-spend-us43500-year-after-school-classes-their

Chen, M.-H., Tseng, W.-T., & Hsiao, T.-Y. (2016). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology*, *49*(1), 69–77. https://doi.org/10.1111/bjet.12526

Chinese Ministry of Education. (2011). 义务教育英语课程标准 (2011版) [English curriculum standards for compulsory education (2011 version)]. Beijing, China: Beijing Normal University Press.

Chiu, Y.-H., Kao, C.-W., & Reynolds, B. L. (2012). The relative effectiveness of digital game-based learning types in English as a foreign language setting: A meta-analysis. *British Journal of Educational Technology*, *43*(4). https://doi.org/10.1111/j.1467-8535.2012.01295.x

Cobb, T., & Horst, M. (2011). Does word coach coach words? *CALICO Journal*, *28*(3), 639–661.

Cook, A. (1995). A review of the IDEA Oral Language Proficiency Test Forms C & D–English. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Dallas, Texas. Retrieved from https://files.eric.ed.gov/fulltext/ED380512.pdf

Csizér, K. (2017). Motivation in the L2 classroom. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 418–432). New York, NY: Routledge.

Dean, C. B., Hubbell, E. R., Pitler, H., & Stone, B. J. (2012). *Classroom instruction that works*. Alexandria, VA: ASCD.

DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge, England: Cambridge University Press.

De Wilde, V., & Eyckmans, J. (2017). Game on! Young leaners' incidental language learning of English prior to instruction. *Studies in Second Language Learning and Teaching, 7*(4), 673–694.

Dörnyei, Z. (2014). Motivation in second language learning. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 518-531). Boston, MA: National Geographic Learning/Cengage Learning.

Doughty, C., & Williams, J. (Eds). (1998). *Focus on form in classroom second language acquisition*. Cambridge, England: Cambridge University Press.

Feng, D. (2006). China's recent curriculum reform: Progress and problems. *Planning and Changing, 37*(1&2), 131–144.

Gass, S. M. (2017). *Input, interaction, and the second language learner* (2nd ed.). New York, NY: Routledge.

Gillon, G. T. (2004). *Phonological awareness: From research to practice*. New York, NY: Guilford Press.

Godwin-Jones, R. (2014). Games in language learning: Opportunities and challenges. *Language Learning & Technology, 18*(2), 9–19. Retrieved from http://llt.msu.edu/issues/june2014/emerging.pdf

Golkova, D., & Hubackova, S. (2014). Productive skills in second language learning. *Procedia–Social and Behavioral Sciences*, *143*, 477–481. https://doi.org/10.1016/j.sbspro.2014.07.520

Heift, T., Mackey, A., & Smith, B. (2019). History, pedagogy, data and new directions: An introduction to the educational technology issue. *Annual Review of Applied Linguistics*, *39*, 1–7.

Henry, A., Davydenko, S., & Dörnyei, Z. (2015). The anatomy of directed motivational currents: Exploring intense and enduring periods of L2 motivation. *The Modern Language Journal*, *99(*2), 329–345.

Hu, G. (2005). English language education in China: Policies, progress, and problems. *Language Policy, 4*(1), 5–24. https://doi.org/10.1007/s10993-004-6561-7

Hung, H.-T., Yang, J. C., Hwang, G.-J., Chu, H.-C., & Wang, C.-C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education*, *126*, 89–104. https://doi.org/10.1016/j.compedu.2018.07.001

Jensen, S. H. (2017). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal*, *34*(1), 1–19. https://doi.org/10.1558/cj.29519

Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Behavioral and Brain Sciences*, *3*(1), 12–19.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education*, *55*(2), 427–443. https://doi.org/10.1016/j.compedu.2010.02.007

Kukulska-Hulme, A., Lee, H., & Norris, L. (2017). Mobile learning revolution: Implications for language pedagogy. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 217–233). Oxford, England: Wiley & Sons.

Lasagabaster, D., Doiz, A., & Sierra, J.M. (Eds.) (2014). *Motivation and foreign language learning: From theory to practice*. Amsterdam, The Netherlands: John Benjamins.

Long, M. H. (1985). A role for instruction in second language acquisition: Task based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modeling and assessing second language acquisition* (pp. 77–99). Clevedon, England: Multilingual Matters.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell Publishing.

Long, M. H. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics*, *36*, 5–33.

Mackey, A. (2012). *Input, interaction and corrective feedback in L2 classrooms*. Oxford, England: Oxford University Press.

Masgoret, A., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, *53*, 123–163. https://doi.org/10.1111/1467-9922.00212

Mercer, S. (2020). *Engaging language learners in contemporary classrooms*. Cambridge, England: Cambridge University Press.

Moeller, A., & Roberts, A. (2013). Keeping it in the target language. In S. Dhonau (Ed.), *Multitasks, multiskills, multiconnections: Selected papers from the 2013 central states conference on the teaching of foreign languages* (pp. 21–38). Eau Clarie, WI: Crown Prints. Retrieved from https://digitalcommons.unl.edu/teachlearnfacpub/178/

Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, *41*(2), 287–311. https://doi.org/10.1017/S0272263118000219

Paulson, L. H. (2004). The development of phonological awareness skills in preschool children: From syllables to phonemes. *Graduate Student Theses, Dissertations, & Professional Papers*, 9522. https://scholarworks.umt.edu/etd/9522

Pienemann, M. (1998). *Language processing and second language development: Processability theory.* Amsterdam, The Netherlands: John Benjamins.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912.

Reinhardt, J. (2017). Digital gaming in L2 teaching and learning. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 217–233). Oxford, England: Wiley & Sons.

Rogers, J. (2017). The spacing effect and its relevance to second language acquisition. *Applied*

Linguistics, *38*(6), 906–911. https://doi.org/10.1093/applin/amw052

Scholz, K. (2017). Encouraging free play: Extramural digital game-based language learning as a complex adaptive system. *CALICO Journal, 34*(1), 39–57.

Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly*, *52*(4), 971–994.

Shrum, J. L., & Glisan, E. W. (2009). *Teacher's handbook: Contextualized language instruction.* Boston, MA: Cengage Learning.

Stansfield, C. W. (1991). Idea Oral Language Proficiency Test. In K. Keyser and R. C. Sweetlands (Eds.), *Test Critiques: Volume VIII*. Austin, TX: Pro-Ed.

Sylvén, L. K., & Sundqvist, P. (2017). Computer-assisted language learning (CALL) in extracurricular/extramural contexts. *CALICO Journal, 34*(1), i–iv.

Tabibian, B., Upadhyay, U., De, A., Zarezade, A., Schölkopf, B., & Gomez-Rodriguez, M. (2019). Enhancing human learning via spaced repetition optimization. *Proceedings of the National Academy of Sciences*. *116*(10), 3988–3933.

Uberman, A. (1998). The use of games: For vocabulary presentation and revision. *English Teaching Forum*, *36*(1), 20.

Wang, X. (2009). Second language theories and their influences on EFL in China. *English Language Teaching*, *2*(4), 149–153. https://doi.org/10.5539/elt.v2n4p149

Wenting, Z. (2019). Children make early start at learning English. *China Daily*. Retrieved from https://www.chinadailyhk.com/articles/141/213/135/1548312114049.html?newsId=60205

Yang, J. C., Lin, D., & Chen, S. (2018). Effects of anxiety levels on learning performance and gaming performance in digital game-based learning. *Journal of Computer Assisted Learning*. *34*(3), 324–334.

Yang, J. C., Quadir, B., & Chen, N.-S. (2015). Effects of the badge mechanism on self-efficacy and learning performance in a game-based English learning environment. *Journal of Educational Computing Research* , *54* (3), 371–394 . https://doi . org/10.1177/0735633115620433

Young, M. F., Slota, S., Cutter, A.B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*. *82*(1), 61–89. https://doi.org/10.3102/0034654312436980

Zhang, D. B. (2012). Chinese primary school English curriculum reform. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 67–83). Dordrecht, The Netherlands: Springer.

Zheng, D., Young, M. F., Brewer, R. A., & Wagner, M. (2009). Attitude and self-efficacy change: English language learning in virtual worlds. *CALICO Journal*, *27*(1), 205–231.

Zhou, S. (2010). Comparing receptive and productive academic vocabulary knowledge of Chinese EFL learners. *Asian Social Science*, *6*(10), 14–19. https://doi.org/10.5539/ass.v6n10p14

Appendix A

Sample IDEA Proficiency Test Items

The examiner places the board shown in Figure A1 in front of the child and says: "I am making up a story about a fun day at the park. Will you help me?" The examiner picks up a picture of a ball and gives it to the child and says: "Put this under the picnic table." The correct response is for the child to respond by placing the ball under the picnic table.

Figure A1. Idea Proficiency Test Storyboard.

Internal English Proficiency Test

Sample *Vocabulary Identification* Item

The examiner plays an audio of the word "Sun." The correct response from the child is to point to the image of the sun.

Figure A2. Images for a Vocabulary Identification item.

Sample *Listening for Meaning* Item

The examiner plays an audio of a sentence that describes a picture, for example, "I see ducks. One, two, three … three ducks!" The correct response is for the child to point to the image of the three ducks. Partial credit is given for a related answer such as the image of three pencils.

Figure A3. Images for a *Listening for Meaning* item.

Sample *Speech Production* Item

The examiner asks in English, "What do you see?" The correct response is for the child to say "rose" or "flower." Partial credit is given for a related answer such as "plant."

Figure A4. Images for a *Speech Production* item.

Sample *Conversation* Item

The examiner asks in English a question such as "What color do you like?" The child is expected to respond in English with an appropriate answer, e.g. "green."

Sample *Pronunciation* Item

The examiner says, “Please listen carefully to each word and repeat it.” The child hears an audio of the object in the photo, e.g., “cookie”, at most 2 times. The child repeats the word that she/he hears.

Figure A5. Image for a *Pronunciation* item.